Project Proposal

Abstract: In recent history the United States has faced an increasingly dire health disaster in the form of mass opioid dependency and overdose. What is the source of this crisis and who is most likely to be affected? This project will use training data from the Washington Post and the CDC (as well as demographic data from the US Census) to quantify the relationship between opioid prescription rate and overdose rate and attempt to create a supervised machine learning model that can predict these rates based on various demographic aspects.

1. What is the problem you want to solve?

* Can we predict prescription rate and overdose rate given various background factors (location, age, sex, race)

1. Who is your client and why do they care about this problem? In other words, what will your client do or decide based on your analysis that they wouldn’t have done otherwise?

* HHS (U.S. Department of Health and Human Services) or other more localized health services could use the information to better target their addiction and recovery programs
* Law enforcement agencies (DEA or local agencies) could use the information to better track illicit drug distributors based on their “clientele”
* AMA (American Medical Association) or private law firms could use the information to pursue medical malpractice suits against doctors whose overprescription of opioids has damaged their communities

1. What data are you using? How will you acquire the data?

* Washington Post Dataset - Pain Pills in the USA: <https://www.kaggle.com/paultimothymooney/pain-pills-in-the-usa>
* CDC Opioid Overdose Deaths Dataset <https://data.world/health/opioid-overdose-deaths/workspace/project-summary?agentid=health&datasetid=opioid-overdose-deaths>
* US Census Data by Characteristic <https://www.census.gov/data/datasets/time-series/demo/popest/2010s-national-detail.html>

1. Outline how you’ll solve this problem
2. Is this a supervised or unsupervised problem?
   1. Supervised - data is already labeled
3. If supervised is it a classification or regression problem?
   1. Regression problem - we’re trying to predict rates
4. What variable is it you are trying to predict? (target variable)
   1. Prescription rates and overdose rates
5. What variables will you use as predictors?
   1. All other features (demographics, etc) used to predict

5. What are your deliverables?

* Code and paper, slide deck possible
* Jupyter notebook (can convert to slide deck)